DOCUMENT RESUME

ED 150 199 TM 006 941

AUTHOR Forbes, Jean W.
TITLE Adapting Title I Evaluation to Local District Needs.
PUB DATE Apr 77
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Compensatory Education Programs; Disadvantaged Youth; Educational Objectives; Elementary Secondary Education; *Evaluation Methods; Evaluation Needs; Evaluators; Federal Programs; *Measurement Goals; Models; *Program Evaluation; *School Districts; State Programs; Testing Problems; Test Selection
IDENTIFIERS *Elementary Secondary Education Act Title I; *Oregon

ABSTRACT

Oregon has recognized the problems involved in the evaluation of Elementary Secondary Education Act Title I programs, and has worked toward solutions, while recognizing the differing needs of local, state, and federal departments of education. In order to provide help to the local school districts with respect to evaluation, a statewide committee was organized. This group developed a set of guidelines to be made available to individual Title I districts within the state. Some of the problems that were identified centered on limited personnel resources in the individual districts, whereas others involving philosophical differences represented by the perceived needs of local districts as compared to state or federal agencies. The committee also explored nine alternative evaluation models, which are summarized. The purposes of evaluation and issues regarding the selection of tests which measure achievement gains are also discussed. (Author/MV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

AERA Session 9.03

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Adapting Title I Evaluation to Local District Needs[1]

Dean W. Forbes
Portland, Oregon, Public Schools (Area 2)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dean W. Forbes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED150199

INTRODUCTION

The outcomes of Title I projects, both at the local and national levels, are presently getting increasingly critical scrutiny. In order to document these outcomes effectively sound evaluation is essential. This presents complications in any region that is characterized by diversity of population or community type. The present comments will be directed toward those circumstances encountered in the state of Oregon and to some of the efforts made to improve evaluation in light of existing situations.

Oregon is characterized by small communities with correspondingly small school districts. Naturally, this provides many small concentrations of pupils. The typical pattern of educational administration is characterized by the community or consolidated school district. (The county district, so common in most of the United States, is almost unknown.) As a result of this, many of the Title I projects in the state are small in size, limited in resources, involve correspondingly small budgets, and maintain a strong desire for local autonomy. This presents serious problems with respect to project evaluation since these smaller districts cannot be expected to have trained and experienced research or evaluation personnel. In far too many portions of the state such personnel are not even available in a supporting capacity from the county intermediate educational district. Over a number of years, it became apparent that some way must be devised to assist the local district with respect to strategies for planning and carrying out Title I project evaluations.

PURPOSES OF EVALUATION

At this point it is necessary to keep in mind the possible purposes that can be served by Title I evaluation. (Since the writer is an evaluator for a local school district, it is understandable that the hierarchy of priorities influencing this presentation will reflect the needs of the local school

TM006 941

---

[1]- Paper presented at the Annual Meeting of the American Educational Research Association, New York City, April 1977

system.) Among these are: (1) to help the local district, and its program, work most effectively with children, to determine how effectively its goals have been met, and to help identify any changes in emphasis necessary for future years; (2) to document to the state the effectiveness of the program, and to provide the state with information which it can accumulate to provide a statewide description; and (3) to provide information that can be transmitted to the federal government (colloquially this is known as "paying the bills").

PROBLEMS AND CONFLICTS INVOLVED IN TITLE I MEASUREMENT

Thus we see that Title I evaluation presents a troublesome paradox. Logic decrees that there must be overall measures of general performance and growth whenever we are looking at the state, regional, or national level (a type of measurement well served by the survey achievement test). This is necessary so that conclusions can be made with respect to general program effectiveness. (The federal government has spent substantial effort and money in this direction and some state departments have moved to a statewide mandate limiting the type of measurement which they consider to be acceptable.) At the same time, while the needs of state and federal governments are being met, it is all too possible to do violence to the needs of local projects for individualized information of the sort most useful to the teachers actually working with specific children.

If one requires that all individual projects, to qualify for funding, must adopt a common and specific set of goals there is a stronger logic in this approach to measurement but it becomes even more difficult for local districts to plan the program most appropriate to their specific situation. On the other hand, to the extent that individual projects are given the option of pursuing goals which have been developed in light of their own unique needs, this becomes an inadequate way of trying to estimate gains of specific children (since <u>their</u> individualized programs may focus on sets of goals not congruent with the overall structure set up by the state). It goes without saying that any measurement, and resulting conclusions, can be generalized only so far as permitted by the particular set of goals which is involved.

The paradox hinges upon the fact that survey measurement provides information best suited to administrative uses. It documents the general flow of progress but is greatly limited in individualized information with respect to specific children. It is, unfortunately, precisely this latter type of information which is seen as necessary by many Title I teachers.

Many programs (and this is especially true in the smaller more intimate programs characteristic of so many Oregon communities where a given school system has relatively few children receiving Title I services) tend to be built upon goals and procedures which focus upon that program which is to be presented to the individual child. The measurement of progress in a situation of this sort may or may not be appropriately served by a year-end survey achievement test, and yet that is the only type of measurement ordinarily mandated in statewide programs.

Clearly, in the minds of many people, educational logic favors an approach to compensatory education which is tailored, to the maximum degree possible, to meet the needs of the local situation and specific children involved. (This point of view has strongly influenced Oregon's approach to Title I evaluation.) Logic states that children must be the focus of compensatory programs and it is only by paying attention to the individual child, the child's particular learning pattern and prior experience, and the child's present needs and capabilities, that justice can be done to the child. This would best be served by helping local districts to set goals and develop that instructional program which is most suited to their children and, by selecting from the various evaluation designs mentioned in the latter part of this paper, come up with a strategy for measuring progress of the individual toward the mastery of specific and customized goals. In addition to its other advantages this approach makes possible the gathering of data from transient students as well as those who arrive late or leave early with respect to a given school year (a type of child who is lost in the course of the conventional pre and post-test design using survey achievement tests).

This puts the individual projects in a dilemma. They must serve the needs of children or of administrative requirements. To qualify for funding, and to justify expenditures of such money, they may be forced to spell out goals that to them are either second choice or purely and simply unacceptable (but which

serve the expedient of qualifying the project for funding). Thus the local district has the choice of: (1) meeting its own needs; (2) meeting the needs expressed in statewide goals; (3) meeting both sets of needs; or (4) attempting some sort of compromise. Clearly, a problem exists.

Taking account of this dilemma, and recognizing the diversity of need and intent from one community to another, the Oregon Title I Evaluator recruited a statewide committee to address the problem. This group, representing a wide variety of background and experience, set out to develop some guidelines which would present strategies and evaluation models that were sufficiently varied that they would meet the needs of virtually any conceivable project, would provide adequate measurement against which to evaluate program effectiveness, and would still be practical for use by personnel in even the smallest school district. (Early in the project it was recognized that this approach to evaluation would limit aggregatability of data, since it is necessary to categorize programs by type and then combine data over similar programs. The benefits to the individual district, especially the smaller ones, were felt to justify this limitation. Later developments suggest procedures which will overcome this difficulty. These will be discussed presently.)

As has just been stated, the committee has worked toward the development of procedures which would make adequate provision for individual local projects. Special attention has been paid to those districts favoring an individualized diagnostic-prescriptive approach to education. Attention has been simultaneously paid to procedures which help the state gather systematic and aggregatible statewide data. This would meet local needs and still provide an overall picture of the effectiveness of state Title I projects.

It is shortsighted to ignore this problem concerning statewide accumulation of evaluation data (especially in light of indications of possible federal mandate with respect to Title I evaluation procedures). Since statewide aggregation of data is most practical with survey achievement testing it is impossible to reject such procedures even though such tests are seen by many persons to have limited relevance to local projects. We seem, in fact, to be forced into option (3) mentioned previously (meeting <u>both</u> sets of needs, local and state). Now we need to develop procedures which will do this with minimum redundancy and inefficiency from the standpoint of the local district

and still provide reliable data to the state Title I office while permitting the major emphasis to be placed on local program needs.

The rationale would, in brief, involve the individual districts getting the information which is useful to them and then supplementing this by broad scale surveys of those achievement areas indicated by a statewide structure of major goals (these invariably involve reading and arithmetic). In this way the state can speak to general development in major skills in relatively common terms while at the same time preserving the tremendous richness of data from individual districts who are pursuing a much more "child oriented" approach to Title I education. In essence, this says that the district can do what it really needs to do and then add on that layer of evaluation needed to provide information which the state requires.

Now, however, another serious problem emerges. The tools which are appropriate to statewide surveys have, to date, invariably been of the sort encountered in the commercially developed proprietary achievement battery (such as included in the list that RMC has recommended). It is an unfortunate fact that such testing is very time consuming. Even a partial battery (reading, language, and arithmetic sub-tests) of such tests can easily involve several hours of testing time. Another problem involves the fact that the only students that will be tested are those that happen to be in attendance at the times of survey testing (which ordinarily are scheduled at the beginning and end of the school year). In many districts this means that easily 20 per cent to 40 per cent of the total number of Title I children are not measured at all. It is very difficult even to estimate the extent of the bias this introduces into test data.

If it so happens that a school district feels that their measurement needs are adequately met by such survey testing, there usually is no problem since it is possible for the district to set up its program in terms of one of the tests from such recommended lists. Districts with well established diagnostic-prescriptive programs ordinarily see no point at all in such testing. In those fairly rare circumstances where districts have sufficient resources that they have developed and standardized tests built around their local curriculum there is a problem since these tests will not have "national" norms and will sometimes vary in content from the typical standardized battery. It is unfortunately

also true that districts with such resources (and such locally developed tests) ordinarily have a sufficiently long span of longitudinal data that they are reluctant or unwilling to change to a commercially published test purely to meet governmental requirements. For such districts the use of such tests requires a double measurement program with resulting inconvenience and loss of class time. To all those projects which "march to a different piper" such testing is merely a bother.

It is clear that in Oregon the required utilization of a survey achievement battery selected from a specified list would present, at best, an inconvenience, and in many cases would be seen as a mere "nuisance" requirement. In addition to this the process of aggregating data from a number of commercial test batteries builds in an accumulation of measurement error which is of concern to some of us despite the much publicized hopes of the Anchor study. For this reason it seems that there is great merit in examining alternatives which can provide common measurement across the state and which would provide estimates of state parameters that are equivalent to those which would result from the use of a test selected from a limited list of pre-identified titles. The ideal procedure would be one which would minimize the amount of additional (or redundant) testing which a district would find necessary to meet state level evaluation needs and which would still keep measurement error to a minimum.

Is there any possibility of identifying a test which could be a statewide requirement for each and every Title I project and which would still be sufficiently brief that it would not impose a major hardship on any district? Such a test would have to be short and still be reliable. It would also have to be relevant to the instructional programs being used (it goes without saying that this last characteristic would team with reliability to provide what we normally consider validity). Furthermore, it would need to provide valid data for a statewide survey of the broad goals involved in Title I.

Examinations of the Mental Measurements Yearbook and Tests in Print identify no probable candidates. Neither does examination of publisher's catalogues give any encouragement. It becomes readily apparent that such a test is not commercially available and would have to be developed. Fortunately, such a test can be developed with minimal effort, and minimal cost. In addition it could be ready for use in a rather short period of time.

POSSIBLE SOLUTIONS TO TITLE I MEASUREMENT PROBLEMS

It takes very little imagination, after careful consideration of the problem, to realize that some of the more recent developments in test theory and test construction offer a solution to the specific problem and, in fact, a resolution of the entire paradox. There have been, for some time, procedures which permit the development of tests which provide large scale achievement surveys to be carried out with extreme parsimony. Until recently they largely have been overlooked. Their use would make it possible for local districts to carry out their preferred measurement programs and still meet state and federal needs with minimal supplementary testing.

As you are all aware, for some years the test calibration procedures first developed by George Rasch, and currently receiving a great deal of development and dissemination by Dr. Ben Wright of the University of Chicago, lead the way to a far greater flexibility in achievement testing than had previously been possible. Most of you will know the underlying rationale and many of you will know specific details of the process so we will not go into those here. Suffice it to say that, using items which have been calibrated with the Rasch procedures, it is entirely possible to develop short tests in reading and mathematics which far exceed in brevity anything commercially available. (Such tests could be on the order of 10 items each in reading and mathematics if the goal was limited to the estimation of statewide, or regional, parameters.)

First of all, before the critics sound off, it must be acknowledged that such a test would not provide adequately reliable measurement for use with an individual child--this is done by the district's own preferred testing program. However, with the numbers of children that are involved in a statewide Title I survey adequate reliability would be ensured so that you would get very accurate estimates of statewide, and even districtwide (in the case of the large to middle-sized districts) parameters. This is precisely what is needed for state and federal purposes.

Such procedures would permit a statewide description of Title I progress and would provide virtually the same degree of measurement accuracy as could be developed by much larger tests (for instance, conventional survey achievement tests). There is the limitation that was mentioned previously. The "full length" test could be used for individual measurement and the 10 item

test could only be used for estimate of large group parameters. For this reason some districts might choose to use such a test to supplement some other survey test they were already using to measure the progress of individual children.

Clearly this is one alternative worthy of exploration. There are still problems, however, and one of the greatest of these centers on the fact that such a test assumes the availability of adequate pools of calibrated items (which presently are almost nonexistent). It also requires access to national norms and these ordinarily are not a part of pools of calibrated items (in most cases, they are not seen as essential or even desirable).

In the last several years the state of Oregon has had a number of school systems that have been very actively and cooperatively involved in validating Rasch logistic procedures for public school achievement testing, and in the development of calibrated pools of test items. This originated in Portland, Oregon, and has centered in the metropolitan area surrounding that city. More recently it has also involved participation by the Oregon State Department of Education and, in the neighboring state of Washington, a number of individual districts plus the State Department of Education. (At the present time the California State Department of Education is also getting involved.) These pools of items would be a promising source of raw material for the development of adequate "quick and efficient" statewide surveys for Title I purposes.

A relatively small amount of research and development would be necessary to accomplish this since procedures for assembling tests in terms of predicted and planned operating characteristics have received a great deal of development by those same districts; the technology and procedures are already known. The actual assembly of short tests, and their field testing and validation, is by no means an excessive job and one involving procedures in which a number of Oregon school districts are well experienced. (Here I will speak only of the Oregon component of the overall movement since the other states are developing their own preferred procedures.)

It seems, then, that the first assumption (the availability of appropriate item files) can be met. The other assumption (compatibility with national

norms) presents a different kind of problem but one that has an equally simple solution. Those who have picked up some experience in the use of calibrated test materials and procedures will immediately see that it is a very practical operation to "link" a logistic scale describing any commercial achievement battery and with that which underlies a file of calibrated items. Once this is done, performance on a test built from calibrated items derived from that file can be mathematically transformed and described in terms of the norms of whichever commercial battery or batteries have been cross linked.

The procedure simply involves administering one or more tests of the battery which is selected as a reference point (the linking must be done separately for each test) and at the same time giving a small selection of items from the calibrated file. It is then a straightforward task to calibrate the battery test and mathematically describe its relationship to the selection of linking items from the calibrated item pool so that a transformation can be made. In this way the metric describing the calibrated pool of items can be expressed in the same terms as the norms of the achievement test.

This could be done with one battery, or several, although this is unimportant because once a cross link is made to one set of commercial, "national", norms it is possible to report the entire state of Oregon in terms of data which is completely compatible with other constituencies using that same commercial battery. To the extent that the "Anchor Test" program was successful such information then can be combined with a much larger data base including all tests included in the Anchor Study.

By these procedures it is entirely possible for the state Title I office to establish a modest amount of mandated testing (something on the order of 30 minutes of testing time) for state purposes, to supplement whatever else the district is using for its own purposes. This would provide comparable data across all children in the state of Oregon who are receiving Title I service (who were available at the times of testing of course) and would thus provide completely aggregatible data across the entire state. Furthermore, this would be completely aggregatible, by the U. S. Office of Education, with data from other states. It would, at the same time, permit the local districts to place the major evaluation emphasis upon their own specific goals and upon customized individualized programs for individual children.

Now that the needs of the state have been met, let's look at some of the specific points that have been addressed by the committee in developing guidelines to help local Title I project personnel plan and carry out the evaluation program best suited to their local needs.

SELECTION OF MEASUREMENT MATERIAL

Since many smaller districts do not even have personnel experienced in the evaluation and selection of measurement tools, and recognizing the fact that no evaluation can be any stronger than the instrumentation used for measurement, a necessary starting point has been the development of guidelines for the selection of tests (or any other measurement materials). Great emphasis has been placed on the necessity of careful test selection utilizing all available information furnished by the publisher and other sources. Particular attention has been given the need for verifying that any test under consideration is appropriate to the specific program for which it is intended. This will be recognized as a practical application of content validity, since if the test will not measure those things which the program has taught it can have no valid use as a measurement tool. Upon occasion it is also necessary to consider predictive validity.

It goes without saying that reliability has also received substantial emphasis since no test can exceed in validity the reliability which it inherently possesses.

Another major point involves consideration of the various available types of testing materials and the selection of that type which will be most appropriate to the specific project (among those discussed are survey achievement tests, diagnostic achievement tests, criterion or goal-referenced mastery tests, and various types of inventories or rating scales necessary for some of the more esoteric goals).

In examing test content it also becomes important to check that there is not built-in bias. Since the state of Oregon has Title I programs with concentrations of various minority groups this is an especially important procedure (different projects around the state have concentrations of Black children, American Indian children, Spanish American children, and children of one or another specific foreign population groups).

It is also essential that consideration be given to the type of norms available for any test under consideration, as well as the time of year when such norms were developed. It is unnecessary to go into any detail here since we all realize that sound measurement requires using tests (if they are to be interpreted in a norm-referenced fashion) that have been normed for the particular times of the year at which measurement is to be carried out.

A final point receiving consideration is practicality (in terms of time required, administration techniques--whether or not professional psychometrists are required, scoring procedures, and all of the many factors which may render an otherwise promising test difficult if not impractical to use).

ACCOMMODATING PROGRAM DIVERSITY

Since different programs in various districts address differing goals we must deal with a number of major types of programs which emerge. Some programs deal with growth throughout an academic year and can nicely be handled by virtually any procedure involving pre and post-testing. They are probably well served by survey achievement tests. Many Title I projects address the remediation of specific weaknesses of individual children. These may involve remediation during an entire school year or until that time when the child reached a preselected criterion level at which time the child can be "mainstreamed" back into the regular program (probably to make way for another child who now has greater need). Programs of this sort obviously must begin with some type of diagnostic testing and probably will utilize some type of instructional regime dealing with sequences of learning packages and mastery tests focused on specific points where remediation is necessary. Some projects are so situated that it is possible, and perhaps desirable, to compare the growth of project children with other (comparable) children not in the Title I program.

This wide variety of local projects, many of which are very small in size, makes it impossible to carry out effective evaluation at the local level by limiting measurement to a required statewide norm-referenced survey achievement test. This prevents reliance on some of the elegant but rather unrealistic models such as those that have been nominated by one or another firm of research consultants.

In an effort to accommodate the wide range of diversity represented by Title I

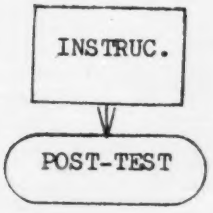
programs in the state of Oregon a variety of evaluation designs have been identified and will briefly be described.

EVALUATION MODELS

The first group of models to be described (1 through 6) are appropriate to long-term projects (either full or half year). The last group of models (7 and 8) will fit the needs of a situation where short-term measurement is necessary. This would include situations where diagnostic-prescriptive programs are being used, as well as those situations where student transiency is sufficiently high that relatively few of the children in a given school in the fall of the year will be there to finish the year in the same location.

1. Measurement of Status at End of Project - This is a "post-test" only design and necessitates comparison to some sort of norm. In most situations this will involve the project group only although it is possible to use this design in situations where one wants to compare the project group with some type of comparison group. It is essential to keep in mind that it is very difficult to get an effective comparison group in most real-life situations and for that reason this is a model of limited utility.

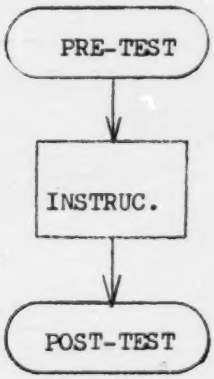
1(a) Post-Test Only



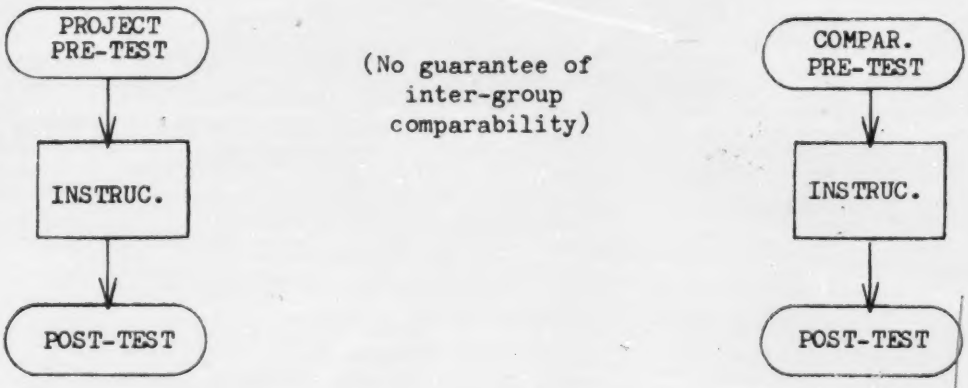
1(b) Post-Test Only (added comparison group)

2. Gain During Duration of Project - This design will involve pre and post-testing and determination of the growth increment that intervenes. It also can be used with a project group only, or with the project group plus a comparison group. Interpretation generally can be made in terms of established norms (and change in group status) although with some types of tests, and in some situations, it might be preferable to test to determine whether or not project group gains had been significantly greater than those of the corresponding comparison group.

2(a) Pre and Post-Test

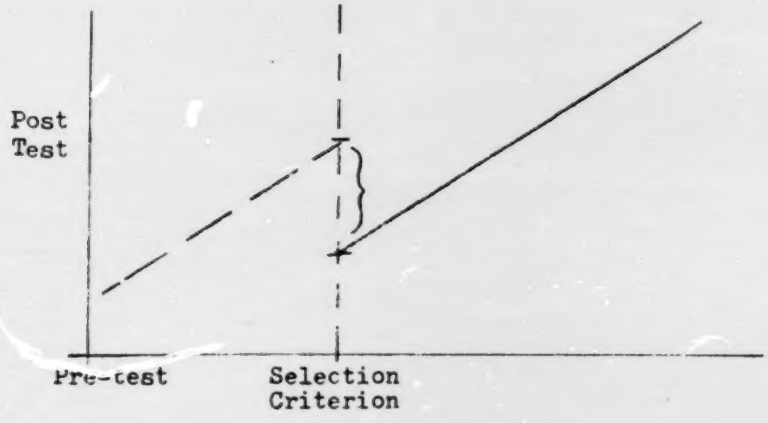


2(b) Pre and Post-Test (added comparison group)

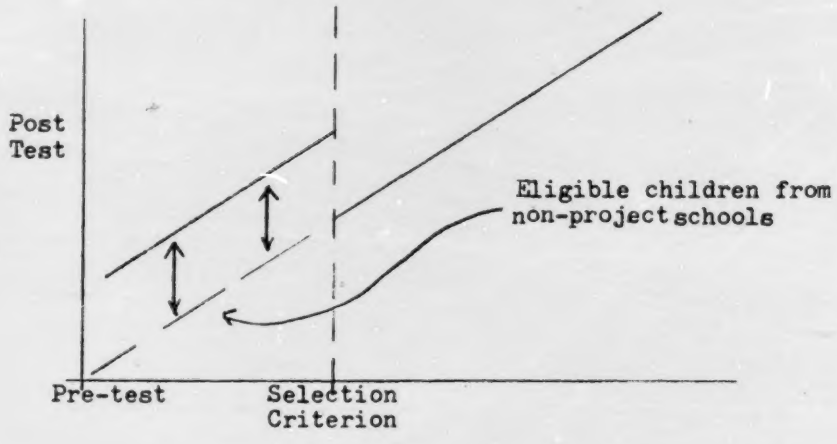


3. Regression of Post-Test on Pre-Test (bivariate) - This model occurs in two variations. 3a is a "pre and post-test" regression discontinuity model such as is described in the various RMC publications. 3b can best be described as a pre-post "parallel" regression model since the non-project population is not truncated below the criterion which identifies project eligibility. This permits tests of linearity with respect to the function being measured in the total population.

3(a) Regression Discontinuity

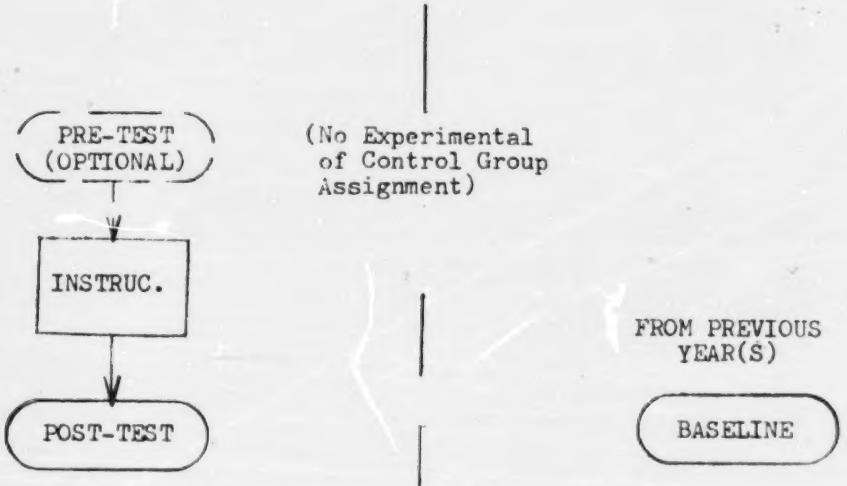


3(b) Parallel Regression



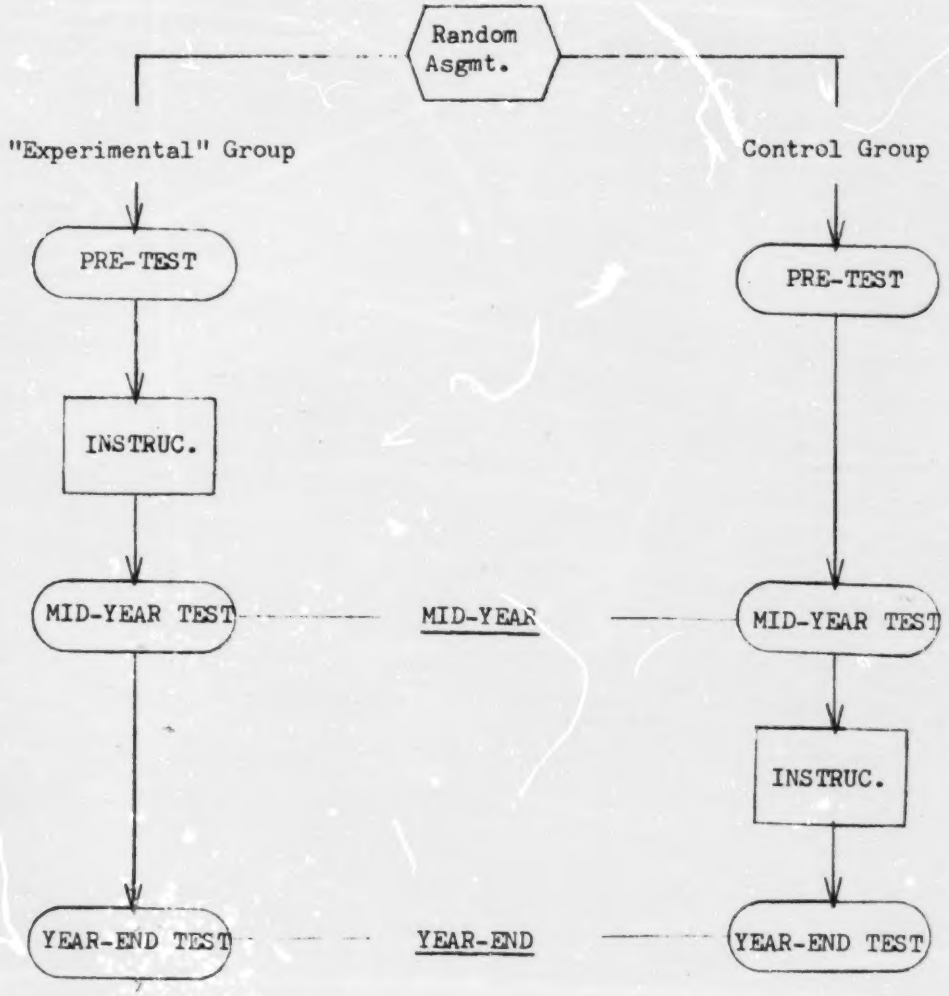
4. Regression of Post-Test on Multiple Predictors - Model 4 is also a regression model but utilizes multiple regression procedures so that prediction, instead of being made by a single variable, is being made by a combination of two or more. 4a is a multiple regression discontinuity model and 4b is a multiple parallel regression model (ordinarily the regression models involve a sufficient degree of background and experience on the part of the evaluator, and sufficient data processing support, that relatively few districts will be in a position to utilize them).

5. Historical Control - This model determines status at the end of a project (although under some circumstances it can measure growth during the course of a project if appropriate pre and post-test preliminary baseline information is available). This is a model appropriate to those situations where it is considered desirable to compare project performance with an adequate comparison group but where a simultaneous comparison group cannot be used (this might occur in a relatively small district where virtually all eligible children are in the project). If population changes are occurring slowly it is entirely possible to utilize data from one or more preceding years to establish a baseline which provides a basis for comparison. It goes without saying that measurement must have been identical in the previous years which are being used as a basis for comparison, and it is obvious that this type of procedure will work only in the initial year or two of Title I involvement; after the passage of one or two project years any suitable pre-project comparison data becomes sufficiently remote in time that its use is very questionable.



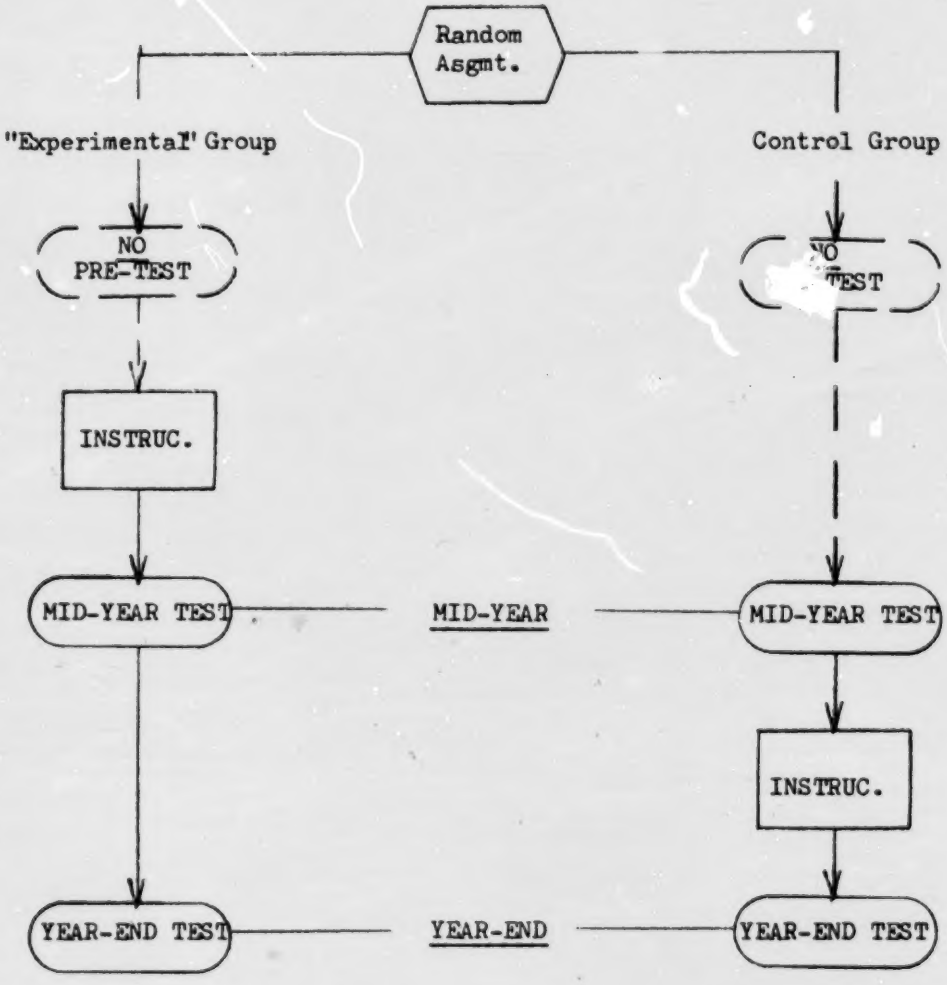
6. Sequential Control - Model 6 can best be described as a sequential control model since it provides a basis for the effective comparison of project participants with a rigorously selected control group. The crucial feature in this model is identification of all project eligible children at the beginning of the year and then their random assignment to two groups. During the first half of the year one group gets treatment while the other group serves as control. At mid-year the status of the two groups is measured and then they are switched so that the control group for the first half of the year now becomes a treatment group during the second half. This has a number of advantages: (1) it is a true experimental design with random assignment to experimental and control groups; (2) it provides strong control group data without denying service to any eligible child; (3) it permits stretching project resources to a larger number of children than might otherwise have been served. (It goes without saying that it is necessary to keep very close track of student mortality in both

6(a) Sequential Control (Pre and Post-Tests)

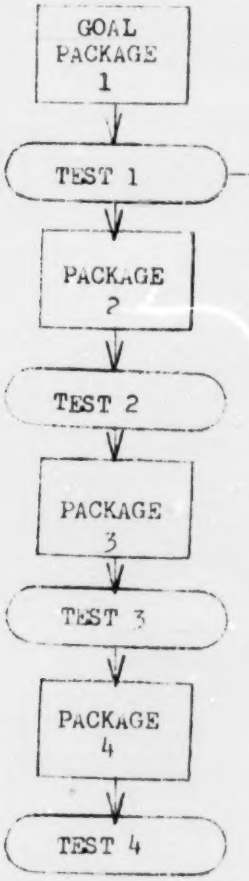


groups since it is entirely possible that movement of children in or out of the program could distort what had started out as acceptably matched groups.) In addition to other advantages, the fact that the first "experimental group" does not receive project service during the second half of the year gives some indication of the durability of any apparent effect which had appeared up to the mid-point when the shift was made. This model can be used with a pre and post-test situation (which would necessitate three testing periods during the year, at the beginning of the year, midway at the switchpoint, and at the end of the year) and is called Model 6a. Model 6b is the identical model eliminating the pre-test at the beginning of the year. This is entirely legitimate so long as the initial placement of children has been truly random although the possibility of datum errors increases as does the movement of children in and out of the program.

6(b) Sequential Control (Post-Test Only)

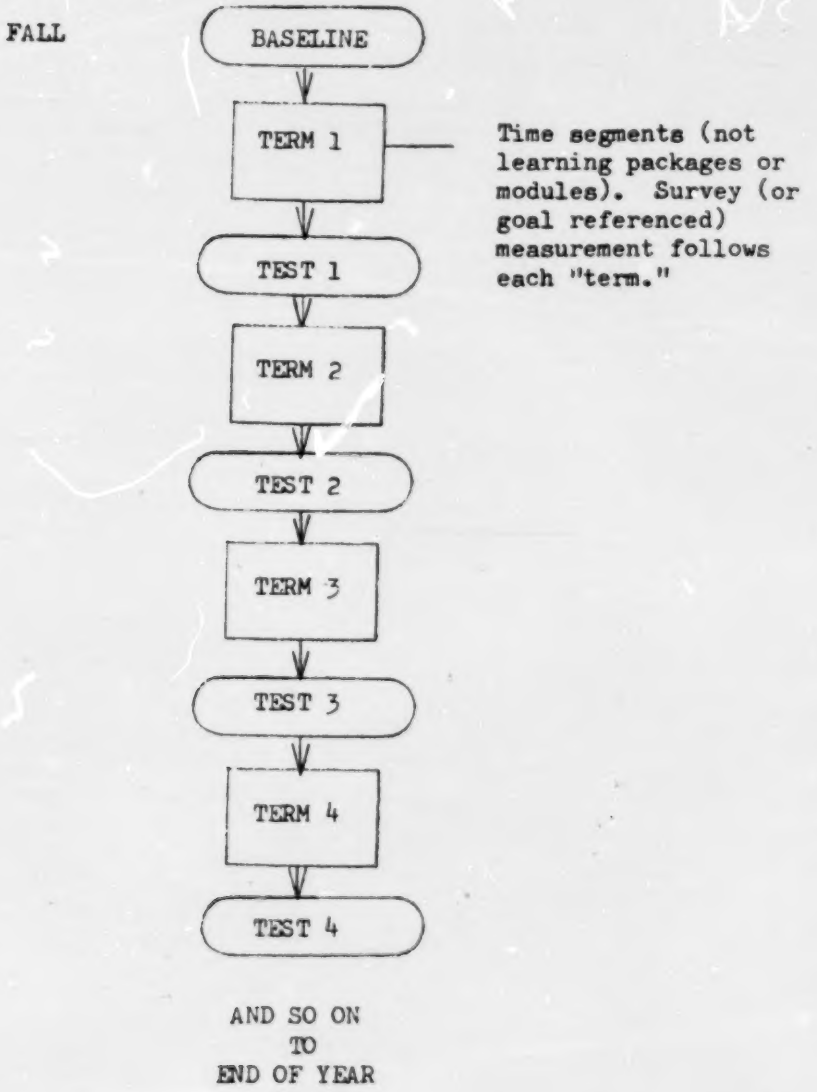


7. Individualized Prescriptive Program Measurement - This model is appropriate where the program consists of a sequence of "learning package and test" components and generally includes all diagnostic and prescriptive programs. These programs are characterized by the situation where a student is presented with a series of learning packages prescribed to meet diagnosed needs and where each learning package will be terminated by a mastery test to determine whether or not the student is ready to move to something else. Measurement with this model will ordinarily be goal referenced or criterion referenced and very little, if any, norm referenced measurement is really suitable (although periodic, or year-end survey testing could be used as a supplement--see Model 9). This model is appropriate to situations where the project priorities involve extreme individualization of service beginning with diagnosis and subsequent prescription. It is very appropriate to those situations where there is a high rate of student transiency.



AND SO ON

8. Periodic Content Referenced Measurement - This ordinarily takes the form of a sequence of short instructional terms each one of which is followed by a test over content which has been covered. This may occur at some regular interval (such as every six weeks) or may be scheduled to coincide with natural segments of the instructional program. This model will salvage some data in those situations where there is such high transiency that few, if any, students beginning the year in a school are still there at its end. The types of tests to be used in a situation of this sort must be selected with extreme care, or carefully developed in light of the local curriculum. (It is entirely possible that this type of situation would be better served by Model 7.)



9. Combination Model - Model 9 uses the strategies of either Model 7 or Model 8 and supplements them by a year-end summary testing (presumably using a survey achievement test). Model 9 can provide useful additional data in those situations where Model 7 has been applied out of philosophical preference. It is less useful where such a decision is one of necessity forced by population movement.

CONCLUDING COMMENTS

Despite this rather narrow, and perhaps unbalanced, presentation it should be emphasized that the Oregon Title I Evaluation Committee has addressed all aspects of the total evaluation process (as the resultant handbook demonstrates). It has carefully considered such topics as: (1) goal setting; (2) development of research questions; and (3) data collection, analysis, and presentation (along with many others). An earlier task force had developed guidelines for the needs analysis phase of Title I planning.

SUMMARY

Oregon has recognized the problems involved in Title I evaluation and data reporting at the local district/project level. Some of the problems centered on limited personnel resources in the individual districts whereas others involve philosophical differences represented by the perceived needs of local districts as compared to state or federal agencies. A statewide committee has addressed these problems and prepared a handbook designed to assist local districts with the evaluation process. It has also explored alternative evaluation models in an effort to reach rapprochement between the data needs of local project staffs and state and federal Title I offices.